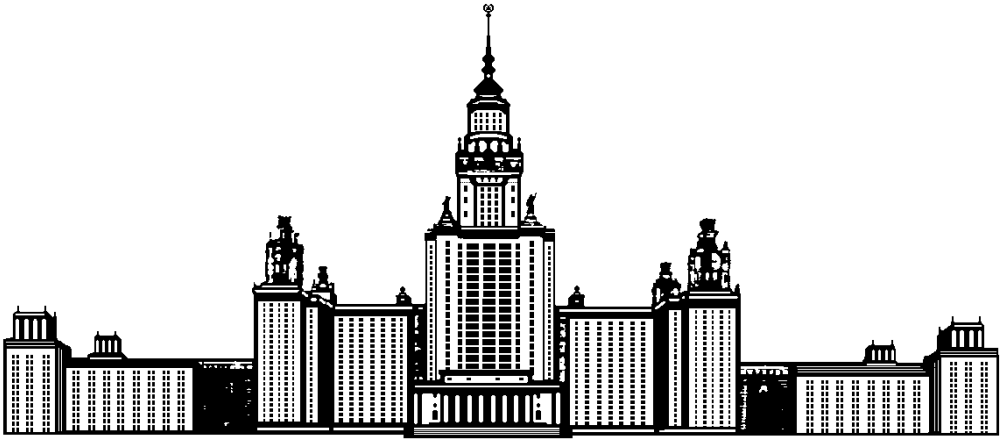
Московский государственный университет
имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования



МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Методы повышения эффективности моделей машинного обучения, основанные на различных принципах снижения размерности

Работу выполнил:
Хомутов Никита Юрьевич

Научный руководитель:
д.ф.-м.н., ведущий н.с. ВЦ РАН
О.В. Сенько

Москва, 2017

Содержание

1 Аннотация

Ансамбли моделей машинного обучения часто используются для повышения обобщающей способности итогового алгоритма, позволяя базовым алгоритмам компенсировать недостатки друг друга. Один из естественных подходов к синтезу ансамбля – использование выпуклых комбинаций результатов предикторов. Наиболее эффективные ансамбли получаются при комбинироваии алгоритмов, основанных на различных принципах.

Линейные модели являются популярным инструментом при анализе данных и разработке моделей машинного обучения, нередко находят свою применимость в случае, когда доступно мало данных.

В данной работе описывается методика построения моделей, использующих избранные оптимальные выпуклые комбинации предикторов, построенных по отдельным признакам изучаемого объекта, для снижения размерности нового признакового описания и учеличения итоговой обобщающей способности модели.

2 Введение

Человечество постоянно сталкивается с задачами классификации, восстановления регрессии. Будь то задача определения систолического давления по кардиограмме, оценки кредитоспособности заёмщика, оценка точки плавления химических соединений в условиях отсутствия точной физически корректной модели или требуемых для симуляции вычислительных ресурсов, разработка системы "свой-чужой".

Сфера применения методов машинного обучения огромна. Существует множество задач, не относимых к "большим данным которые, тем не менее, имеют актуальность. При разработке методов медицинской диагностики исследователи нередко сталкиваются небольшим количеством пациентов, множеством признаков и сложными физиологическими процессами, связывающие наблюдаемые признаки с исследуемым явлением. Распространённые причины малочисленности данных – дороговизна их извлечения, редкость исследуемых событий.

В сфере "малых данных"особенно остро встаёт вопрос переобучения модели, эффективной валидации результатов. Это создаёт спрос на простые и робастные методы, налагающие меньшие требования на данные. Это объясняет популярность и эффективность линейных моделей.

Имея набор уже построенных моделей, можно построить новую путём построения ансамбля моделей. Использование ансамблей позволяет моделям компенсировать ошибки друг друга, давая более качественную итоговую модель. Существуют различные методики построения ансамблей, и конструирование выпуклой комбинации результатов имеющихся алгоримов – одна из них. Выпуклые комбинации наиболее эффективны в случае использования моделей, основанные на различных принципах, обученных на различных подвыборках и подмножествах признаков.

В работе [1] представлена новая линейная модель, объединяющая простоту линейных моделей и эффективность выпуклых комбинаций. Производится построение простых линейных моделей построенных на отдельных признаках, после чего производится поиск их оптимальной

выпуклой комбинации. Процесс поиска представляет собой комбинаторную задачу, в процессе которой исследуются выпуклые комбинации различных подмножеств индивидуальных линейных регрессий. Возникла идея использовать не только самую лучшую оптимальную комбинацию, но и другие исследуемые комбинации, построенные на других подмножествах признаков, выделяя различные паттерны в данных. Таким образом, вместо одного итогового результата мы получаем множество других, которые можем воспринимать как новые производные признаки в исследемой задаче. Встала задача отбора исследуемых комбинаций, схожая с задачей отбора признаков: исспользуя результаты относительно обобщающей способности ансамблей, основанных на выпуклых комбинациях, в результате отбора находились значительно качественные алгоритмы. После чего уже возможно применение стандартных моделей машинного обучения в новом признаковом пространстве.

Экспериментально была показана эффективность разработанного метода в ряде прикладных задач, наблюдается улучшение по сравнению с моделью ElasticNet. Наблюдается эффективность схемы отбора выпуклых комбинаций по сравнению с наивным подходом "лучшие k комбинаций". Метод показал эффективность в условиях малого количества наблюдаемых объектов, зашумлённости данных.

В дальнейшем, в соответствии с принципами глубинного обучения, можно исследовать многоуровненые модели, использующие метод построения выпуклых комбинаций в качестве линейных преобразованый между слоями модели. Такой подход позволит строить многоуровненые представления данных, выделяя различные паттерны в них, при этом накладывая на данные меньше ограничений, чем того требуют современные нейронные сети. Если будет показана эффективность такого подхода, то это существенно расширит круг задач для методов глубинного обучения.

2.1 Постановка задачи

В рамках данной работы поставлены следущие задачи:

- Реализовать алгоритм нахождения оптимальных выпуклых комбинаций
- Разработать методику преобразования признакового пространства с помощью выпуклых комбинаций
- Разработать метод отбора комбинаций с целью увеличения разнообразия представленный в ансамбле комбинаций для уменьшения итоговой ошибки
- Экспериментально оценить эффективность метода в регрессионных задачах, сравнить с исходным признаковым пространством на модели ElascticNet.

3 Методы валидации моделей

Модель, обученная на тщательно собранной выборке может давать на другой выборке существенно более худшие результаты, чем на известной выборке. Ухудшение результатов при применении параметрической модели для новых данных возникает из-за избыточной сложности пространства параметров. Лишние степени свободы "тратятся"на подгонку под конкретную обучающую выборку, выделяя зависимости, которых нет во всей задаче. Из-за этой подгонки под лишние зависимости получаем смещённую оценку на параметры модели, что приводит к ухудшению качества в целом. Проблема переобучения статистических моделей – одна из ключевых проблем в машинном обучении.

Для борьбы с переобучением применяются различные методы [3]

- Минимизация теоретической оценки качества модели
- Накладывание дополнительных ограничений на пространство параметров модели с целью уменьшить количество степеней свободы – регуляризация
- Минимизация оценки качества на скользящем контроле

Процедура скользящего контроля заключается в следующем. Пусть $S = \{s_n | 1 \leqslant n \leqslant\}$ – данная нам выборка; μ – метод обучения, строящий алгоритм по выборке; а $Q(a, \hat{S})$ – функционал качества, определяющий качество алгоритма a на выборке $\hat{S}$.

Выборка S разбивается N различными способами на две непересекающиеся подвыборки $S = S_n^{train} \cup S_n^{test}$, здесь n – номер разбиения. Для каждого разбиения на обучающей подвыборке строится алгоритм $a_n = \mu(S_n^{train})$ и вычисляется значение функционала качества на тестовой подвыборке $Q_n = Q(a_n, S_n^{test})$.

Среднее арифметическое значений Q_n по всем разбиениям называет-

ся оценкой скользящего контроля:

$$CV(\mu, S) = \frac{1}{N} \sum_{n=1}^{N} Q(\mu(S_n^{train}), S_n^{test})$$

Распространены различные стратегии построения разбиений исходной выборки на непересекающиеся подвыборки.

- Полный скользящий контроль: оценка строится по всем $C_{|S|}^k$ разбиениям исходной выборке, при которых под тестовую подвыборку выделяется k объектов. В частном случае при $k = 1$ получаем контроль по отдельным объектам (leave-one-out CV). Практический интерес представляет именно случай $k = 1$, так как при $k \geqslant 2$ задача становится вычислительно сложной. Однако, случай $k \geqslant 2$ может быть интересен для теоретических исследований и в тех редких ситуациях, когда есть возможность вывести эффективную вычислительную формулу.

- Контроль на отложенных данных: оценка строится по одному случайному разбиению $N = 1$. Среди существенных недостатков данного метода: уменьшение длины обучающей подвыборки приводит к смещённой оценки (завышенной, пессимистичной) вероятности ошибки; существенная зависимость оценки от разбиения; высокая дисперсия оценки.

- Контроль по отдельным объектам. Случай полного скользяшего контроля при $k = 1$, соответственно, $N = |S|$ – строится $|S|$ разбиений, при которых один объект составляет тестовую подвыборку, а остальные – обучающую $S = (S \setminus \{s_n\}) \cup \{s_n\}$. В [2] показано, что при некоторых условиях контроль по отдельным объектам является асимптотически оптимальным.

  Процедура имеет ряд преимуществ: каждый объект в контроле участвует ровно один раз, а длина обучающей выборки меньше

лишь на единицу длины полной выборки, что уменьшает смещённость оценки. Недостаток же в том, что процедура обучения запускается $|S|$ раз, что сказывается на вычислительной ресурсоёмкости метода. Но для задач с малым количеством объектов в наблюдаемой выборке это не критично.

Стоит отметить, что распространены функционалы качества Q, сравнивающих векторы ответов алгоритмов на тестовой подвыборке с эталонными ответами $Q = F(y_{predict}, y_{test})$. Нетривиальные примеры таких функционалов: ROC-AUC, коэффициент детерминации. В таких случаях целесообразно не усреднять значения Q_n по N разбиениям, а вычислять функцинал на скомбинированных векторах ответов:

$$LOO(\mu, S) = F(y^{all}_{predict}, y^{all}_{test})$$

Здесь $(y^{all}_{predict})_n = a_n(s_n)$, $y^{all}_{test} = y$.

- Контроль по q блокам (q-fold CV). Выборка случайным образом разбивается на подвыборки равной (или почти равной) длины: $S = S_1 \cup \cdots \cup S_q$. Затем рассматривается q разбиений, в которых одно из подмножеств используется как тестовая подвыборка, а остальные – для обучения $S = (S \setminus S_q) \cup S_q$. Данный метод позволяет уменьшить вычислительную сложность оценки по сравнению с контролем по отдельным объектам, так как обучение производится q раз вместо $|S|$.

- Контроль по $r \times q$ блокам. В данном методе процедура контроля по q-блокам повторяется r раз. Метод обладает всеми преимуществами метода контроля по q-блокам.

4 Ансамбли алгоритмов

Использование ансамблей алгоритмов может существенно улучшить качество итоговой модели [6]. Комбинирование моделей позволяет компенсировать ошибки одних алгоритмов за счёт других. Различные алгоритмы $A_1, \ldots, A_r$ могут быть получены как путём применения различных методов прогнозирования или распознования, как путём обучения на различных подвыборках, так и путём выделения различных подмножеств признаков.

Можно было бы отобрать один наилучший алгоритм среди заданных, однако, исползование всех алгоритмов, как правило, обладают более высокой обобщающей способностью, чем отдельные базовые модели.

Композиция не знает внутреннего устройства базовых алгоритмов. Для неё базовые алгоритмы – чёрные ящики, позволяющие только обучаться на заданной выборке и вычислять ответ для заданного объекта.

Существуют различные подходы к построению ансамблей алгоримтов. Основные принципы построения ансамблей алгоритмов:

- Специализация. Признаковое пространство объектов разделяется на области, в каждой из которых строится собственный алгоритм, решающий задачу только для рассматриваемой области. Эксплуатируя принцип "разделяй и влавствуй исходная задача разбивается на более простые подзадачи.

- Агрегирование. В этом подходе не учитывается наблюдаемая информация об объекте, не происходит специализации делегирования. Используются только результаты, выданные базовыми алгоритмами. К подобным методам относятся комитеты большинства, бэггинг, бустинг, выпуклые комбинации.

При этом, распространённые способы построения ансамблей:

- Итеративно. Базовые алгоритмы строятся поочерёдно, при этом каждый следующий алгоритм старается компенсировать недостатки предыдущих. Данная стратегия – жадная, и она не гарантирует

построения наилучшей композиции. Тем не менее, такой подход нашёл своё место в практическом применении. Ярчайшим примером такой стратегии является бустинг.

- Параллельно. Базовые алгоритмы настраиваются независимо друг от друга. При этом стараются выбирать такие условия построения алгоритмов, что бы получать не похожие друг на друга модели. Примеры данного подхода: бэггинг, метод случайных подпространств, генетические алгоритмы. Так же сюда можно отнести и выпуклые комбинации (хотя, с помощью выпуклых комбинаций можно перевзвесить набор алгоритмов, полученный с помощью итеративных методов).

- Глобальная оптимизация всех базовых алгоритмов – тяжёлая задача, требует знания внутреннего устройства алгоритмов. Пример – EM-алгоритм разделения смеси распределений.

4.1 Выпуклые комбинации

Выпуклые комбинации – подход к построению ансамблей с помощью агрегирования заданных алгоримов. Алгоритмам назначаются неотрицательные веса, с которыми они линейно входят в композицию. Комбинация представляет собой "центр масс"алгоритмов с учётов их весов.

Рассмотрим задачу восстановления регрессии, в строим алгоритмы, вычисляющие прогноз переменной Y по описанию объектов. Рассмотрим множество прогнозирующих алгоритмов $A_1, \ldots, A_r$.

Пусть f_i – прогноз, вычисляемый алгоритмом A_i. Тогда

$$\Delta_i = \mathbb{E}_\Omega (Y - f_i)^2$$

является математическим ожиданием квадрата ошибки прогнозирования для алгоритма A_i. Здесь Ω – исследуемое распределение, из которого и получена обучающая выборка. Введём обозначение расстояния между

прогнозами ρ_{ij} – математического ожидания квадрата отклонения друг от друга прогнозов, вычисляемых алгоритмами A_i и A_j. То есть

$$\rho_{ij} = \mathbb{E}_\Omega \left((f_i - \mathbb{E}_\Omega f_i) - (f_j - \mathbb{E}_\Omega f_j) \right)^2 .$$

Пусть $c_1, \ldots, c_r$ – неотрицательные коэффициенты выпуклой комбинации такие, что

$$\sum_{i=1}^{r} c_i = 1.$$

Обозначим через $\hat{f}$ выпуклую комбинацию прогнозов, вычисляемых алгоритмами ансамбля $A_1, \ldots, A_r$. То есть

$$\hat{f} = \sum_{i=1}^{r} c_i f_i$$

Для ошибки выпуклой комбинации справедливо следующее выражение:

$$
\begin{aligned}
\hat{\Delta} =&\mathbb{E}_\Omega(Y-\hat{f})^2 = \mathbb{E}_\Omega(Y-\sum_{i=1}^{r} c_i f_i)^2 = \\
=&\mathbb{E}_\Omega\left(\sum_{i=1}^{r} c_i(Y-f_i)\right)^2 = \\
=&\mathbb{E}_\Omega\left(\sum_{i=1}^{r} c_i(Y-f_i)\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \left((\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)^2 - (\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)^2\right) = \\
=&\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega((f_i-Y)(f_j-Y)) + \\
&+\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \left(\mathbb{E}_\Omega(f_i(\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j))) - \mathbb{E}_\Omega(f_j(\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j))) - (\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)^2\right) = \\
=&\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega\left(2(f_i-Y)(f_j-Y) + (f_i-f_j)(\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j) - (\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)^2\right) \\
=&\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega(Y-f_i)^2 - \frac{2}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega(f_i-Y)^2 - \\
&-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega\left(-2(f_i-Y)(f_j-Y) - (f_i-f_j)(\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j) + (\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)^2\right) = \\
=&\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega(Y-f_i)^2 - \\
&-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega\left((f_i-Y) - (f_j-Y) - (\mathbb{E}_\Omega f_i - \mathbb{E}_\Omega f_j)\right)^2 = \\
=&\sum_{i=1}^{r} c_i \mathbb{E}_\Omega(Y-f_i)^2 - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \mathbb{E}_\Omega\left((f_i-\mathbb{E}_\Omega f_i) - (f_j-\mathbb{E}_\Omega f_j)\right)^2 = \\
=&\sum_{i=1}^{r} c_i \Delta_i - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \rho_{ij}
\end{aligned}
\tag{1}
$$

Итого,

$$
\hat{\Delta} = \mathbb{E}_\Omega(Y-\hat{f})^2 = \sum_{i=1}^{r} c_i \Delta_i - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j \rho_{ij} \tag{2}
$$

Принимая во внимание, что все квадратичные отклонения ρ_{ij} всегда

неотрицательны, а коэффициенты выпуклой комбинации $c_1, \ldots, c_r$ неотрицательны, получаем неравенство

$$\hat{\Delta} \leqslant \sum_{i=1}^{r} c_i \Delta_i$$

Соответственно, вооружившись формулой ошибки выпуклой комбинации приходим к выводу что использование качественных разнородных комбинаций (с высоким расстоянием ρ_{ij}) приводит к увеличению отрицательного члена в (2), уменьшая итоговую ошибку ансамбля.

Этим объясняется, почему для конструирования ансамблей алгоритмов стараются использовать алгоритмы, основанные на разнообразных моделях машинного обучения, обучают модели на подвыборках и подмножествах переменных.

5 Линейные модели

Рассмотрим стандартную задачу многомерного регрессионного анализа. Переменная отклика Y предсказывается по n переменным-признакам $X_1, \ldots, X_n$ с помощью линейной регрессионной функции

$$\beta_0 + \sum_{i=1}^{n} \beta_i X_i$$

Предполагается, что вектор регрессионных коэффициентов $\beta = (\beta_0, \ldots, \beta_n)$ вычисляется по обучающей выборке $\{(y_i, x_{1j}, \ldots, x_{nj}) | j \in 1, \ldots, m\}$.

Популярные методы, используемые для поиска регресионных коэффициентов линейных моделей в задачах высокой размерности – гребневая регрессия, Лассо [4] и эластичная сеть [5]. Эти методы основаны на решении регуляризованной задачи минимизации средней квадратичной ошибки, и задачи имеют вид:

$$\min_{\beta \in \mathbb{R}} \left\{ \sum_{j=1}^{n} \left(y_i - \beta_0 - \sum_{i=1}^{n} \beta_i x_{ij} \right)^2 + \lambda P(\beta) \right\}$$

где $\lambda \geqslant 0$ и $P(\beta) = P(\beta_0, \beta_1, \ldots, \beta_n)$ – функция штрафа. Для метода Лассо функция штрафа имеет вид $\sum_{i=1}^{n} |\beta_i|$, для гребневой регрессии – $\sum_{i=1}^{n} \beta_i^2$, для эластичной сети –

$$(1 - \alpha) \sum_{i=1}^{n} \beta_i^2 + \alpha \sum_{i=1}^{n} |\beta_i|$$

здесь $\alpha \in [0, 1]$ – параметр, характеризующий компромис между регуляризацией по методу Лассо и гребневой регрессии. Метод, используемый в модели эластичной сети, на практике показал свою чрезвычайную эффективность на множестве задач, что, однако, не гарантирует его исключительную эффективность.

6 Модель выпуклой регрессии

Отметим существование другого способа введения регуляризации линейной модели, основанного на решении оптимизационной задачи с ограничениями:

$$\begin{gathered}\min_{\beta\in\mathbb{R}}\left\{\frac{1}{m}\sum_{j=1}^{m}\left(y_i-\beta_0-\sum_{i=1}^{n}\beta_i x_{ij}\right)^2\right\}\\ C_1(\beta_0,\ldots,\beta_n)\geqslant 0\\ \ldots\\ C_k(\beta_0,\ldots,\beta_n)\geqslant 0\end{gathered}\tag{3}$$

В работе [1] вводится модель, удовлетворяющая ограничениям 3, называемая выпуклой регрессией.

Показывается, что следующая процедура решает задачу 3:

1. Построение одномерных линейных регрессий $R_1,\ldots,R_n$. R_i, обученных стандартным методом наименьших квадратов на выборке (Y,X_i), то есть только на одном признаке:

$$R_i=\beta_0^{ui}+\beta_1^{ui}X_i$$

2. Построение оптимальной выпуклой комбинации одномерных линейных регрессий $R_1,\ldots,R_n$, максимально коррелирующей с переменной отклика Y:

$$\tilde{R}(\mathbf{c})=\sum_{i=1}^{n}c_iR_i\qquad K(\tilde{R}(\mathbf{c}),Y)\to\max_{\mathbf{c}}\tag{4}$$

3. Построение одномерной линейной регресси, выполняющей масштабирование и сдвиг оптимальной выпуклой комбинации:

$$\bar{R}=\beta_0^c+\beta_1^c\tilde{R}(\mathbf{c_{opt}})$$

Описанная выше процедура и называется процедурой выпуклой регрессии.

В работе доказывается ряд утверждений, показывающие эквивалентность решений 3 и решений выпуклой регрессии.

6.1 Структура пространства выпуклых комбинаций индивидуальных линейных регрессий

Самая вычислительно тяжёлая часть – поиск оптимальной выпуклой комбинации – решение задачи 4. Для решения данной задачи теоретически было исследовано пространство выпуклых комбинаций индивидуальных линейных моделей.

Рассмотрим ансамбль предикторов $\mathbf{r} = \{r_1, \ldots, r_l\} \subseteq \{R_1, \ldots, R_n\}$, состоящий из некоторых индивидуальных линейных моделей. Здесь $l = |\mathbf{r}|$ – количество элементов в ансамбле.

Каждому ансамблю можно поставить множество выпуклых комбинаций

$$\sum_{i=1}^{l} c_i r_i \qquad \sum_{i=1}^{l} c_i = 1 \qquad \forall 1 \leq i \leq l : c_i > 0$$

Таким образом, сопоставляются те и только те выпуклые комбинации, в которых предикторы из $\mathbf{r}$ имеют положительные веса, а остальные – нулевые.

Обозначим

$$\hat{D}_l = \{\mathbf{c} | \sum_{i=1}^{l} c_i, c_i \geq 0, i = 1, \ldots, l\}$$

– замкнутый l-мерный симплекс

$$D_l = \{\mathbf{c} | \sum_{i=1}^{l} c_i, c_i > 0, i = 1, \ldots, l\}$$

– открытый l-мерный симплекс

Таким образом, в соответствие ансамблю предикторов $\mathbf{r}$ можем поставить выпуклые комбинации с весами из открытого l-мерного симплекса D_l. Обозначим за $P(\mathbf{r}, \mathbf{c})$ – соответстствующую выпуклую комбинацию:

$$P(\mathbf{r}, \mathbf{c}) = \sum_{i=1}^{l} c_i r_i$$

Определение Ансамбль предикторов $\mathbf{r}$ называется несократимым относительно коэффициента корреляции, если

- либо ансамбль состоит из единственного предиктора $l = 1$ и при этом предиктор имеет положительную корреляцию относительно переменной отклика $K(Y, r_l) > 0$
- либо ансамбль состоит из нескольких предикторов $l > 1$ и удаление любого набора предикторов из комбинации приводит к ухудшению максимальной корреляции между переменной отклика Y и соответствующими выпуклыми комбинациями. То есть существует такой вектор $\mathbf{c}^* \in D_l$, что $\forall \mathbf{c}' \in \hat{D}_l \setminus D_l$ выполняется

$$K[Y, P(\mathbf{r}, \mathbf{c}^*)] > K[Y, P(\mathbf{r}, \mathbf{c}')]$$

Определение Несократимый ансамбль предикторов $\mathbf{r}$, включающий l предикторов называется нерасширяемым ансамблем, если не существует такого несократимого ансамбля с числом предикторов $l + 1$, включающего каждый предиктор из $\mathbf{r}$.

Для поиска оптимальной выпуклой комбинации важно учитывать следующее:

- Был разработан тест несократимости ансамбля предикторов $\mathbf{r}$
- Кроме того, ансамбль $\mathbf{r}'$, не являющийся несократимым, не может быть подмножеством какого-либо несократимого ансамбля $\mathbf{r}^*$

Таким образом, получаем алгоритм поиска несократимых ансамблей предикторов:

После нахождения желаемых ансамблей алгоритмов находим соответствующую оптимальную выпуклую комбинацию с максимальным коэффициентом корреляции к переменной отклика Y. Процедура поиска соответствующей точки на открытом симплексе D_l так же описана в [1].

Исходные параметры: $\{R_1, \ldots, R_n\}, Y$

Результат: incompressible combinations

Проверить множества, состоящие только из одного предиктора, на несжимаемость ;

Проверить множества, состоящие только из пар предикторов (валидных на предыдущем шаге), на несжимаемость ;

$valid\_pairs :=$ валидные пары ;

$combos\_prev\_level := valid\_pairs$;

$level := 3$;

до тех пор, пока $combos\_prev\_level \neq \varnothing$ **выполнять**

- Выдать $combos\_prev\_level$;
- $combos\_curr\_level := \varnothing$;
- **цикл** $combo \in combos\_prev\_level$ **выполнять**
  - $last\_pred := combo[last\_added]$;
  - **цикл** $new\_pred \in valid\_pairs[last\_pred]$ **выполнять**
    - **если** $combo \cup \{new\_pred\}$ – *несжимаемая* **тогда**
      - $combos\_curr\_level$.append$(combo \cup \{new_pred\})$
    - **конец условия**
  - **конец цикла**
- **конец цикла**
- $combos\_prev\_level := combos\_curr\_level$

конец цикла

Алгоритм 1: Алгоритм поиска несократимых ансамблей предикторов

7 Использование выпуклых комбинаций для геренации нового признакового пространства

Промежуточные ансамбли предикторов, вычисленные в процессе решения задачи (4) (и запуска алгоритма 1), могут использовтся в качестве новых признаков. Такие ансамбли являются несжимаемыми и используют все представленные в них признаки. По каждому такому ансамблю генерируем соответствующую оптимальную выпуклую комбинацию.

Таким образом мы приходим к проблеме генерации информативных признаков, к той же проблеме, что встречается в парадигме глубинного обучения. Предполагается, что использование выпуклых комбинаций позволит выделять высокоуровненые паттерны в данных, сохраняя свойства выпуклых комбинаций относительно обобщающей способности ансамбля.

Пусть у нас есть выпуклые комбинации $\mathbf{C} = \{C_1, \ldots, C_m\}$. Являясь алгоритмами, геренирующими ответ по признаковому описанию объектов выборки, мы можем получить ответы $\mathbf{Z} = \{Z_1, \ldots, Z_m\}$, которые используем в качестве новых признаков объектов.

Однако же существует проблема: среди срегенированных выпуклых комбинаций есть много, настроенных на похожих множествах предикторов. Встаёт вопрос: как отобрать k новых признаков.

Рассмотрим ошибку выпуклой комбиации подмножества новых признаков $\mathbf{Z}' \subset \mathbf{Z}$ длины $k = |\mathbf{Z}'|$ при предположении, что все новые признакаи входят с одинаковым весом (мы не делаем на данном этапе никаких априорных предпочтений):

$$\frac{1}{k}\sum_{Z\in\mathbf{Z}'}\Delta(Z) - \frac{1}{2k^2}\sum_{Z_1\in\mathbf{Z}', Z_2\in\mathbf{Z}'}\rho(Z_1, Z_2) \to \min_{\mathbf{Z}'\in\mathbf{Z}, |\mathbf{Z}'|=k} \tag{5}$$

Наш метод отбора комбинаций заключается в решении задачи 5. Таким образом, выбираем подмножество признаков, чья выпуклая ошибка минимизируется при предположении равенства весов.

8 Эксперименты

Для проверки эффективности метода перехода в новое признаковое пространство, составленное из выпуклых комбинаций, в регрессионных задачах выполнялась следующая экспериментальная схема. В связи с малочисленностью данных валидация результатов происходила в режиме контроля по отдельным объектам (leave-one-out). В качестве метрик качества использовались:

- Коэффициент корреляции к переменной отклика
- Коэффициент детерминации r^2-score.

Строились выпуклые комбинации, отбирались лучшие k из них с учётом 5. Формировалось новое признаковое пространство *combo-k*. Исходное признаковое пространство будем обозначать за *raw*. Так же рассматривалось и признаковое пространство-смесь – объединение *combo-k* и *raw* – *mix-k*.

Экспериментально исследовалось качество моделей машинного обучения, обученных в различных признаковых пространствах: исходном; с использованием отобранных выпуклых комбинаций; смеси исходных признаков и отобранных выпуклых комбинаций. Эксперименты показывают, что в большинстве случаев модель эластичной сети в новом признаковом пространстве имеет лучшую обобщающую способность, а в некоторых случаях значительно лучшую. Так же были и отрицательные результаты, которые свидетельствуют о том, что метод всё же не универсален.

Для различных реальных практических задач были проведены серии экспериментов согласно схеме, описанной выше. Одна серия экспериментов включает в себя оценку качества модели ElasticNet в режиме контроля по отдельным объектам при значении гиперпараметра $\alpha = 0.5$ и $\alpha = 0.99$, отвечающего за соотношение участия $L1$- и $L2$-регуляризаторов. Наилучшая из данных оценок качества воспринимается как базовая.

Затем (в режиме контроля по отдельным объектам) запускается процедура поиска выпуклых комбинаций. Данная процедура имеет дополнительный параметр, позволяющий уменьшить количество рассматриваемых комбинаций, называемый далее $generation\_threshold$. Текущая реализация процедуры предполагает, что сначала рассматриваются комбинации длиной 1, затем комбинации длиной 2, 3 и так далее. Данные шаги поиска комбинации будем называть *уровнями*. При рассмотрении всех комбинаций k-го уровня выбирается комбинация с наибольшим значением коэффициента корреляции к переменной отклика Y, а значение коэффициента корреляции обозначаем как $best\_k$. Тогда при рассмотрении комбинаций $(k+1)$-го уровня мы отбрасываем комбинации, для которых значение коэффициента корреляции с переменной отклика Y меньше, чем:

$$generation\_threshold \cdot best\_k$$

Это дополнительная разреживающая процедура отсекает заведомо низкокачественные комбинации. Параметр $generation\_threshold$ так же может варьироваться в серии экспериментов. Однако, почти во всех сериях используется значение:

$$generation\_threshold = 0.99$$

как эмпирически оптимальное. Такое соотношение позволяет рассматривать комбинации $(k+1)$-го уровня лучшие, так и чуть менее качественные, чем $best\_k$, с рассчётом на численную неустойчивость, а так же для внесения разнообразия в набор рассматриваемых комбинаций.

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.547726515385	0.813129897687
raw	-	0.99	0.545829354141	0.821984006866
combo-5	0.99	0.5	0.560717137768	0.85659597182
combo-10	0.99	0.5	0.676320943425	0.869017346636
combo-30	0.99	0.5	0.733659487326	0.863741259573
combo-30	0.99	0.99	0.74980150554	0.866438781279
combo-45	0.99	0.5	0.736264512325	0.861273472654
combo-60	0.99	0.5	0.738651012427	0.861345690978
combo-60	0.99	0.99	0.754587533735	0.86874001091
combo-100	0.99	0.5	0.736934328828	0.859194469067
mix-5	0.99	0.5	0.698702940213	0.853673986801
mix-10	0.99	0.99	0.754328648312	0.870641973092
mix-15	0.99	0.5	0.753820083523	0.870456411978
mix-30	0.99	0.99	0.755628413529	0.870153034634
mix-60	0.98	0.99	0.759963817137	0.87201814802
mix-100	0.99	0.99	0.773976072117	0.880290397463
mix-120 *	0.99	0.99	0.775583491292	0.881218649188
mix-150	0.99	0.99	0.771012774399	0.878938325088

Таблица 1: Результаты эксперимента для задачи AHal3-MP при пороге генерации 0.99

8.1 Результаты экспериментов

8.1.1 Задача AHal3 Melting Point

В рамках данной задаче требовалось определить точку плавления различных сплавов по их формуле [7] [8]. Из формулы химического вещества извлекались признаки составляющих элементов (56 штук). В выборке представлено 155 объектов.

Две серии экспериментов отличаются параметром *generation_threshold*. Сравнивая результаты экспериментов, приведённых в таб. 1) и таб. 1, не

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.547726515385	0.813129897687
raw	-	0.99	0.545829354141	0.821984006866
combo-5	0.98	0.5	0.560717137768	0.85659597182
combo-10	0.98	0.5	0.676320943425	0.869017346636
combo-30	0.98	0.5	0.733659487326	0.863741259573
combo-30	0.98	0.99	0.74980150554	0.866438781279
combo-45	0.98	0.5	0.736264512325	0.861273472654
combo-60	0.98	0.5	0.738651012427	0.861345690978
combo-60	0.98	0.99	0.754587533735	0.86874001091
combo-100	0.98	0.5	0.736934328828	0.859194469067
mix-5	0.98	0.5	0.698702940213	0.853673986801
mix-10	0.98	0.99	0.754328648312	0.870641973092
mix-15	0.98	0.5	0.753820083523	0.870456411978
mix-30	0.98	0.99	0.755628413529	0.870153034634
mix-60	0.98	0.99	0.759963817137	0.87201814802
mix-100	0.98	0.99	0.773976072117	0.880290397463
mix-120 *	0.98	0.99	0.775583491292	0.881218649188
mix-150	0.98	0.99	0.771012774399	0.878938325088

Таблица 2: Результаты эксперимента для задачи AHal3-MP при пороге генерации 0.98

заметна разница в качестве получаемых моделей. Следовательно, комбинации, добавленные при смене значения порога генерации с 0.99 на 0.98 не участвовали в итоговых моделях.

Во всех экспериментах с переходом в новое признаковое пространство (combo-k и mix-k) наблюдается возрастание качества получаемых моделей. Наилучший результат показала модель, использующая исходные признаки вместе с 120 комбинациями.

8.1.2 Задача AHal3 H

В рамках данной задаче требовалось определить изменение энтальпии величины при некотором термодинамическом процессе. Из формулы химического вещества извлекались признаки составляющих элементов (312 штук). В выборке представлено 158 объектов.

В данной серии экспериментов 3 базовое значение качества было преодалено не во всех случаях перехода в новое признаковое пространство по метрике $r^2 - score$. Однако, при использовании 30 и более комбинаций наблюдается улучшение. При этом значение корреляции при переходе в новое признаковое пространтсво выше, чем базовое, во всех случаях, кроме модели $mix - 5$.

8.1.3 Задача AT_NR2

В данной задаче требовалось спрогнозировать содержание антител к белку NR2 в организме человека, по различных биологическим показателям: содержанию микроэлементов в волосах, возрасту и полу [9] [10]. Выборка представлена 88 объектами в 41-мерном признаковом пространстве.

Результаты, полученные в 4 примечательны тем, что модель эластичной сети показывает низкое качество по метрике $r^2 - score$ на исходный признаках. Качество моделей $mix - 10$ и $mix - 60$ по метрике $r^2 - score$ было ниже нуля в режиме контроля по отдельным объектам. $combo - k$ модели по коэффициенту корреляции с переменной отклика имеют более

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw	-	0.5	0.919156573166	0.959382178849
raw *	-	0.99	0.938290813069	0.968799569799
combo-5	0.99	0.5	0.902328937214	0.973852678677
combo-10	0.99	0.5	0.93163970148	0.972747249364
combo-30	0.99	0.5	0.943901606547	0.972674475036
combo-30 *	0.99	0.99	0.94421283505	0.972528385994
combo-45	0.99	0.5	0.94320285252	0.972577711502
combo-60	0.99	0.5	0.942212372876	0.97236471882
combo-100	0.99	0.5	0.940816316909	0.97106402121
mix-5	0.99	0.5	0.928771683801	0.964419186438
mix-10	0.99	0.99	0.94285758279	0.97315611157
mix-15	0.99	0.5	0.942903690081	0.972568379057
mix-30	0.99	0.99	0.943547042392	0.97305774793
mix-60 *	0.99	0.99	0.945226986452	0.973248342882

Таблица 3: Результаты эксперимента для задачи AHal3-H при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.0675962090177	0.294838607056
raw *	-	0.99	0.0478602276937	0.310067564633
combo-5	0.99	0.5	0.049820327183	0.340170603862
combo-10	0.99	0.5	0.0679992521231	0.361149592342
combo-30	0.99	0.5	0.0801339580406	0.3953423694
combo-30	0.99	0.99	0.0621714889164	0.382474592169
combo-60	0.99	0.5	0.0990769145324	0.385334795287
combo-100 *	0.99	0.5	0.104175154451	0.383966889153
mix-5	0.99	0.5	0.0997246572813	0.350708870014
mix-5	0.99	0.99	0.0147635070406	0.225077732399
mix-10	0.99	0.99	-0.0101050382105	0.222862034979
mix-15	0.99	0.99	0.00565196497931	0.234012337448
mix-60	0.99	0.99	-0.0314006636119	0.126294427171

Таблица 4: Результаты эксперимента для задачи AT_NR2 при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.93568462464	0.968111884625
raw	-	0.99	0.935039284492	0.968045910612
combo-5	0.99	0.5	0.942728718023	0.974270340485
combo-10	0.99	0.5	0.945288795642	0.974210281045
combo-30	0.99	0.5	0.946262420001	0.973861115421
combo-30	0.99	0.99	0.945706576242	0.973233377336
combo-45 *	0.99	0.5	0.947531273248	0.974301490006
combo-60	0.99	0.5	0.947226775039	0.974103982405
combo-100	0.99	0.5	0.945463898142	0.973118173017
mix-5 *	0.99	0.5	0.951813342382	0.975683906575
mix-10	0.99	0.99	0.941969293698	0.970605717431
mix-15	0.99	0.5	0.945236386329	0.972250187172
mix-30	0.99	0.99	0.945593787074	0.972429881675
mix-60	0.99	0.99	0.946624025502	0.972956305994

Таблица 5: Результаты эксперимента для задачи AB2X4 при пороге генерации 0.99

высокое качество, чем *raw*-модели. По метрике $r^2 - score$ так же удалось достичь улучшения для $combo - k$-моделей.

8.1.4 Задача AB2X4

Данная задача так же является химической. В рамках данной задаче требовалось определить параметр A кристаллической решётки для различных сплавов. Из формулы химического вещества извлекались признаки составляющих элементов (95 штук). В выборке представлено 87 объектов.

Исходчя из результатов в таб. 5, наблюдаем улучшение качества по метрикам $r^2 - score$ и коэффициенту корреляции к переменной отклика Y. Наилучший результат был получен на модели $mix - 5$, где к исход-

ным признакам были добавлены 5 комбинаций, отобранных при решении задачи (5).

8.1.5 Задача Chalcopyrite Bangdap

В данной задаче требовалось спрогнозировать величину запрещённой зоны для различных химических веществ. Из формулы химического вещества извлекались признаки составляющих элементов (701 штук). В выборке представлено 133 объекта.

Так же был дополнительно запущен эксперимент по построению модели $combo - 200$, в котором не проводилась процедура, при решении задачи (5) выбора оптимального набора комбинаций. Выбирались 200 комбинаций с наиболее высоким значением корреляции к переменной отклика Y. Обозначим данную модель как $combo - 200(naive)$. В таблице 7 есть результаты данного эксперимента с сравнении с моделью $combo - 200$, где используется введённая ранее процедура отбора комбинаций. Как видим из результатов, введение процедуры отбора было осмысленно и дало существенный прирост как по метрике $r^2 - score$, так и по значению коэффициента корреляции с переменной отклика Y.

Из результатов таб. 6 видно, что так же удалось достичь улучшения по сравнению с базовыми raw-моделями.

8.1.6 Задача Peredat3

В данной задаче требуется определить значение систолического давления по спектру передаточной фукнции кадриограммы-фотоплетизмограммы пациента. В выборке 346 объектов, 61 признак.

В результатах из таб 8 особенно примечательно, что модель ElascitNet на исходных признаках выдавала антикоррелирующие результаты и отрицательные значения около нуля по метрике $r^2 - score$. Переход в новое признаковое пространство существенно улучшил картину. Наилучший результат выдала модель, использующая 5 комбинаций.

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.710756511457	0.849807243853
raw	-	0.99	0.638963967865	0.815011098637
combo-5	0.99	0.5	0.696011725744	0.834437641061
combo-10	0.99	0.5	0.502166283736	0.738922005332
combo-30	0.99	0.5	0.689467109066	0.831280457723
combo-30	0.99	0.99	0.672942475222	0.827237814307
combo-45 *	0.99	0.5	0.704077823867	0.839410409259
combo-60	0.99	0.5	0.720409259365	0.848790772243
combo-60	0.99	0.99	0.682179895755	0.830368160057
combo-100	0.99	0.5	0.727200632674	0.852813084368
combo-100	0.99	0.99	0.669479760587	0.82205384163
combo-200	0.99	0.5	0.735215481178	0.858667060638
combo-250	0.99	0.5	0.735215481178	0.858667060638
combo-300	0.99	0.5	0.732645221004	0.858322596457
mix-5 *	0.99	0.5	0.951813342382	0.975683906575
mix-10	0.99	0.99	0.941969293698	0.970605717431
mix-15	0.99	0.5	0.945236386329	0.972250187172
mix-30	0.99	0.99	0.945593787074	0.972429881675
mix-60	0.99	0.99	0.946624025502	0.972956305994

Таблица 6: Результаты эксперимента для задачи Chalcopyrite Bangdap при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
combo-200	0.99	0.5	0.735215481178	0.858667060638
combo-200 (naive)	0.99	0.5	0.627041553582	0.800875428527

Таблица 7: Сравнение схем отбора комбинаций для задачи Chalcopyrite Bangdap при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	-0.00554865280769	-0.934386399672
raw	-	0.99	-1.0	-0.00580550304558
combo-5 *	0.99	0.5	0.329980393556	0.574712007323
combo-10	0.99	0.5	0.326188657916	0.571519813266
combo-30	0.99	0.5	0.299299102145	0.555993399006
combo-30	0.99	0.99	0.27817620384	0.55840576965
combo-60	0.99	0.5	0.232554706266	0.539576280081
mix-5	0.99	0.5	0.0443820947133	0.525434275482
mix-5	0.99	0.99	0.205662201615	0.553620582531

Таблица 8: Результаты эксперимента для задачи Peredat3 при пороге генерации 0.99

8.1.7 Задача Peredat9

В данной задаче требуется определить значение систолического давления по спектру передаточной фукнции кадриограммы-фотоплетизмограммы пациента. В отличие от задачи Peredat3, использовались другие методы построения признаков. В выборке 433 объекта, 91 признак.

Здесь (в таб. 9) так же удалось достичь улучшения по сравнению с базовой моделью.

8.1.8 Задача SBPGer

В данной задаче требовалось определить систолическое давление по различным клинико-лабораторным данным пациентов НКЦ "Геронтологии-[11]. В выборке 236 объектов, 100 признаков.

Результаты в таб. 10 так же говорят о том, что использование $combo-k$ и $mix-k$ моделей вместо базовых raw дало существенный прирост качества. Лучгий результат дали модели $combo-5$ и $mix-5$.

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw *	-	0.5	0.211664618935	0.489383434225
raw	-	0.99	0.191858670414	0.502433013185
combo-5	0.99	0.5	0.290802684146	0.539964345892
combo-10	0.99	0.5	0.282017086955	0.53159618065
combo-30	0.99	0.5	0.279496566079	0.529148145037
combo-30 *	0.99	0.99	0.292659551899	0.541749107325
combo-60	0.99	0.5	0.2780843225	0.527712215633
mix-5	0.99	0.5	0.239820623196	0.514592270806

Таблица 9: Результаты эксперимента для задачи Peredat9 при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw	-	0.5	-0.556145273676	0.0713009028952
raw *	-	0.99	0.148305657266	0.385157597941
combo-5 *	0.99	0.5	0.181443993272	0.436148624381
combo-10	0.99	0.5	0.174924068019	0.42875230408
combo-30	0.99	0.5	0.171695008029	0.422815725001
combo-30 *	0.99	0.99	0.170928256699	0.425593554474
combo-60	0.99	0.5	0.178458270269	0.428487162775
combo-100	0.99	0.5	0.177449351529	0.427142945531
mix-5	0.99	0.5	-0.209339888078	0.156044187002
mix-5 *	0.99	0.99	0.183237819859	0.43098204292

Таблица 10: Результаты эксперимента для задачи SBPGer при пороге генерации 0.99

пространство	generation_threshold	ElasticNet α	r2_score	корреляция
raw	-	0.5	0.240085855222	0.514654742895
raw *	-	0.99	0.251887336715	0.529530591608
combo-5 *	0.99	0.5	0.283335990448	0.53249452727
combo-10	0.99	0.5	0.284242584629	0.533359947229
combo-30	0.99	0.5	0.288006038222	0.537277894201
combo-30	0.99	0.99	0.276381304259	0.526421435613
combo-60 *	0.99	0.5	0.288640446844	0.538205626312
mix-5	0.99	0.5	0.250874834353	0.522078685871

Таблица 11: Результаты эксперимента для задачи SBPGer при пороге генерации 0.99

8.1.9 Задача Spectr1

В рамках данной задачи так же требуется прогнозировать величину систолического давления. Признаковое пространство формируется на основе спектральных свойст кардиограммы пациента. В выборке представлено 429 объектов, 122 признака.

Результаты в таб. 11 показывают улучшение качества в случае $combo-k$ моделей. Удалось добиться существенного прироста качества на модели $combo-60$ по сравнению с базовыми.

8.1.10 Синтетическая задача gtest

Рассматриваемая задача является синтетической. В данной задаче представлено 550 признаков, только 5% из них релевантны. Переменная отклика линейно зависит от релевантных признаков. Было сгенерировано 100 пар выборок (обучающих и тестовых), в каждой выборке по 40 объектов.

В рамках эксперимента для каждой пары выборок (обучающая и тестовая) выполнялась следующая процедура. Сначала обучалась модель ElasticNet на обучающей выборке, качество оценивалось на тестовой

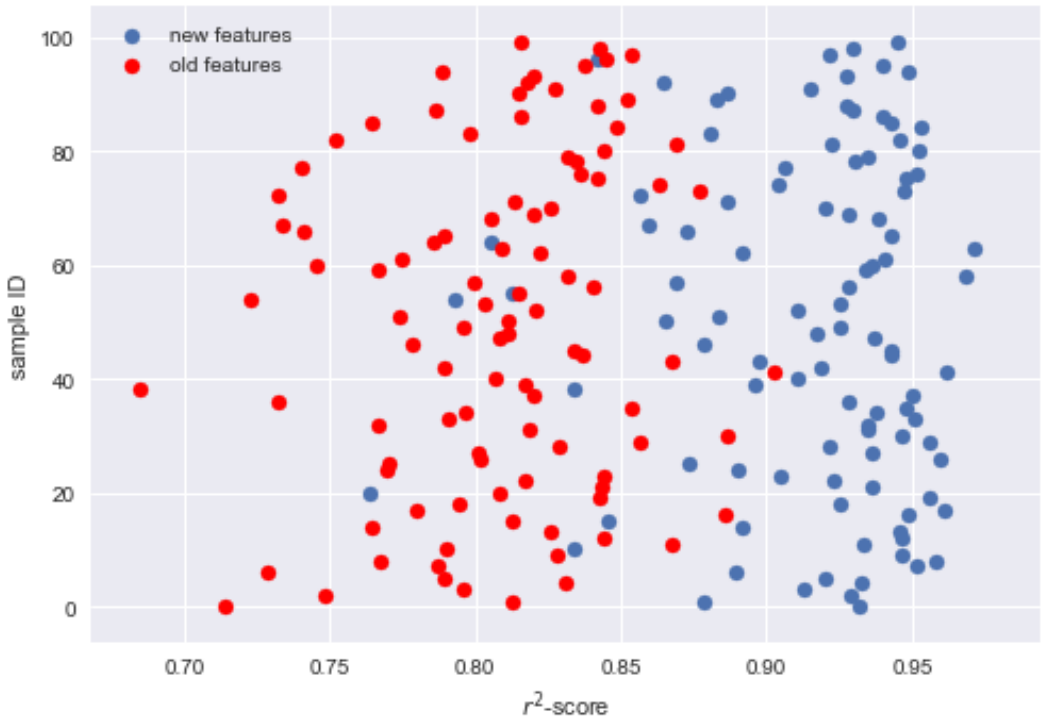


Рис. 1: Изменение качества для различных выборок при переходе в признаковое пространство, составленное из комбинаций.

(красные точки). Затем на обучающей выборке запускалась процедура построения выпуклых комбинаций, отбирались комбинации, коэффициент корреляции с переменной отклика которых был не меньше, чем 95% от величины корреляции для самой лучшей комбинации.

$$corr \geqslant 0.95 \cdot best\_corr$$

Отобранные комбинации использовались в качестве нового признакового пространства, в котором уже производилось измерение качества модели ElasticNet (при обучении на обучающей выборки, тестировании на тестовой).

Как видно из рис. 1, в большинстве случаев наблюдаем заметное улучшение качества для данной синтетической задачи.

9 Заключение

- Разработана методика преобразования исходного признакового пространства с помощью выпуклых комбинаций
- Разработан метод отбора комбинаций с целью увеличения разнообразия представленных в ансамбле комбинаций для уменьшения итоговой ошибки
- Экспериментально показана эффективность метода в сравнении с исходным признаковым пространством на модели ElascticNet.

Список литературы

[1] Regression model based on convex combinations best correlated with response. Dokukin, A.A. & Senko, O.V. Comput. Math. and Math. Phys. (2015) 55: 526. doi:10.1134/S0965542515030045

[2] Li, K.-C. 1987. Asymptotic Optimality for Cp, CL, Cross-Validation, and Generalized Cross-Validation: Discrete Index Set. The Annals of Statistics, 15: 958–975.

[3] Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов. — Математические вопросы кибернетики / Под ред. О. Б. Лупанов. — М.: Физматлит, 2004. — Т. 13. — С. 5–36.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. 58, 267–288 (1996).

[5] H. Zou, T. Hastie, B. Efron, and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Stat. Soc. 67 (2), 301–320 (2005).

[6] Combining Pattern Classifiers: Methods and Algorithms. Kuncheva, Ludmila I. Wiley-Interscience, 2004.

[7] Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химиии материаловедения. – Успехи химии, 2010.

[8] Киселева Н.Н. Н.Д. Ващенко, В.П. Гладун, С.Р. ЛеКлэр, А.Г. Джексон, Прогнозирование неорганических соединений, перспективных для поиска новых электрооптических материалов., 2012.

[9] Клименко Л.Л., Скальный А.В., Турна А.А., Кузнецова А.В., Сенько О.В., Баскаков И.С., Буданова М.Н., Савостина М.С., Мазилина А.Н. Роль селена в многофакторном этиопатогенезе ишемического инсульта..// Микроэлементы в медицине. Т.16, №4, с 28-35

[10] О. В. Сенько , А. М. Морозов , А. В. Кузнецова , Л. Л. Клименко Оценка эффекта множественного тестирования в методе оптимальных достоверных разбиений..//Машинное обучение и анализ данных. 2016, т.2, № 1

[11] А.В. Кузнецова, И.В. Костомарова, Н.Н. Водолагина, Н.А. Малыгина, О.В. Сенько. Изучение влияния клинико-генетических факторов на течение дисциркуляторной энцефалопатии с использованием методов распознавания // Матем. биолог. и биоинформ., 2011, том 6, выпуск 1, страницы 115–146.